26 ℓ)

DOCKET NO. BABU 1-10-42   **PATENT**

#3
660803
S-I

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:

    Shivnath Babu, et al.

Serial No.:   10/033,199

Filed: December 28, 2001

For: SYSTEM AND METHOD FOR COMPRESSING
    A DATA TABLE USING MODELS

Group: 2641

Examiner: N/A

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Alexandria, VA 22313, on May 20, 2003 (Date)
Stephanie Pritt
(Printed or typed name of person signing the certificate)
Stephanie Pritt
(Signature of the person signing the certificate)

Sir:

LETTER TO OFFICIAL DRAFTSMAN

Transmitted herewith are eight sheets of formal drawings to be substituted for the informal

drawings initially filed in the above-identified application for patent.

Respectfully submitted,

HITT GAINES & BOISBRUN, P.C.

David H. Hitt
Registration No. 33,182

Date: MAY 19, 2003
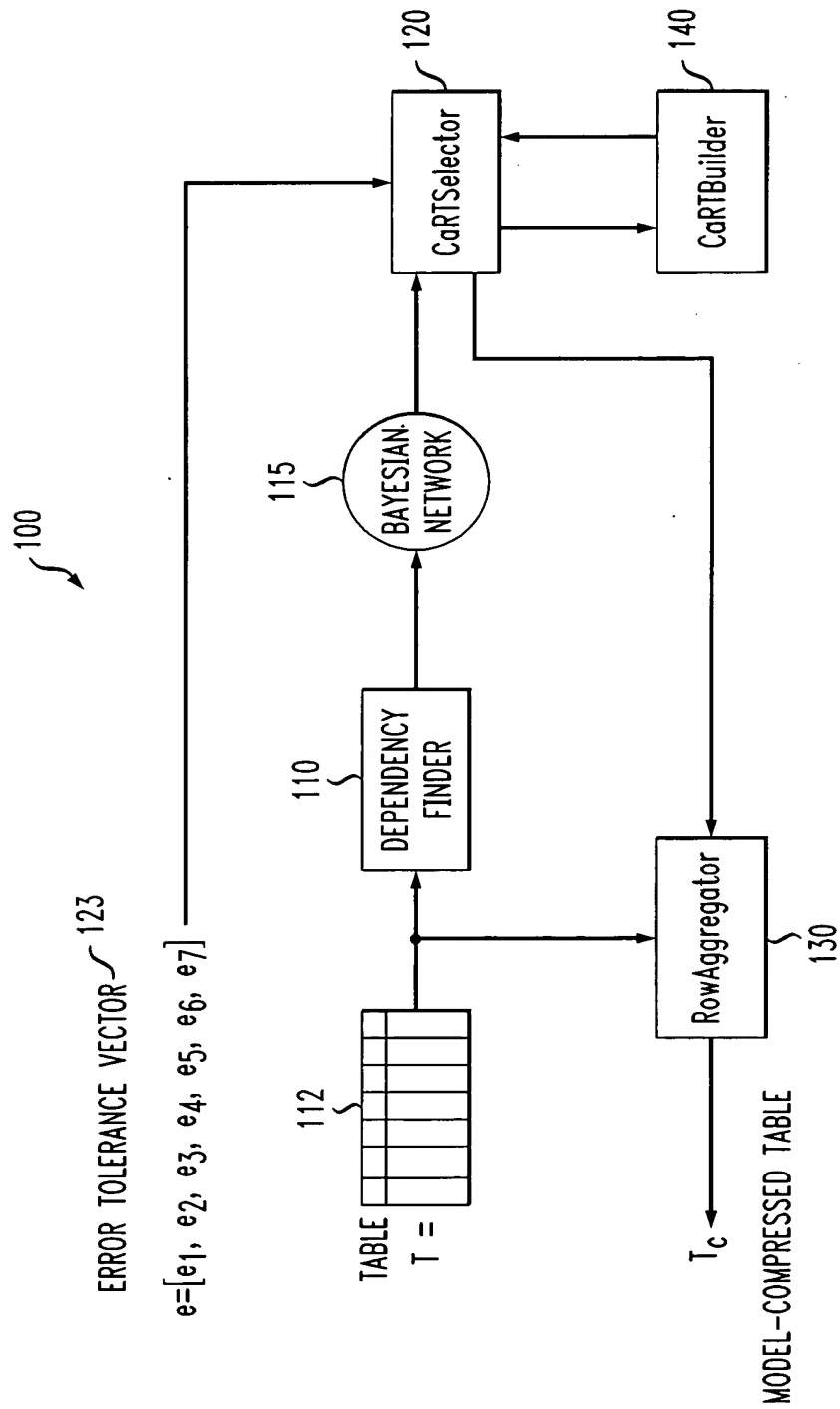
P.O. Box 832570
Richardson, Texas 75083
(972) 480-8800

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

1/8

FIG. 1

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

2/8

FIG. 2

The Greedy CaRT Selection Algorithm

procedure Greedy $(T(X)$, $\bar{e}$, G, θ)

Input: n-attribute table T and n-vector of error tolerances $\bar{e}$;
Bayesian network G on the set of attributes X and
threshold θ on the relative benefit for selecting a
CaRT predictor.

Output: A set of materialized (predicted) attributes X_{mat} (X_{pred}
$= X - X_{mat}$) and a CaRT predictor for each $X_i \in X_{pred}$.

begin

1. $X_{mat} := X_{pred} := \Phi$

2. let $< X_1, X_2,...,X_n >$ be the attributes in X sorted in *topological order* of G

3. for i $:=1,...,n$

4. if $\Pi (X_i) = \Phi$ then $X_{mat} := X_{mat} \cup \{X_i\}$ /* X_i must be materialized if it has no parents in G */

5. else

6. $M := \text{BuildCaRT} (X_{mat} \to X_i , e_i)$

7. if $(\text{MaterCost} (X_i) / \text{PredCost} (X_{mat} \to X_i) > \theta)$ then $X_{pred} := X_{pred} \cup \{X_i\}$

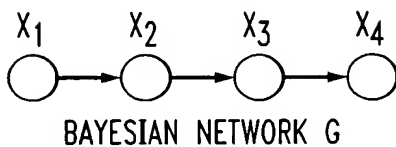8. else $X_{mat} := X_{mat} \cup \{X_i\}$

9. end

10. end

end

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

3/8

FIG. 3A

$$X_1 \quad X_2 \quad X_3 \quad X_4$$

BAYESIAN NETWORK G

FIG. 3B

$$X_1 \quad X_2 \quad X_3 \quad X_4$$

-infty 50 110 50

(b) Gtemp (FIRST ITERATION)

FIG. 3C

$$X_1 \quad X_2 \quad X_4$$

-infty -15 45

Gtemp (SECOND ITERATION)

FIG. 3D

$$X_1 \quad X_2$$

-infty -60

Gtemp (THIRD ITERATION)

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

4/8
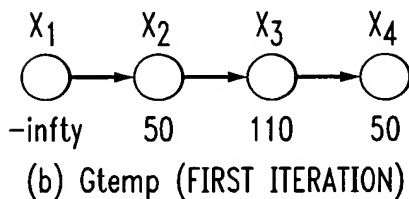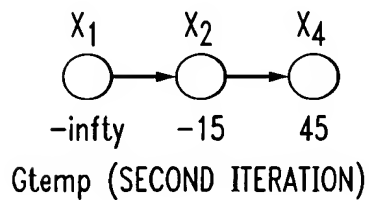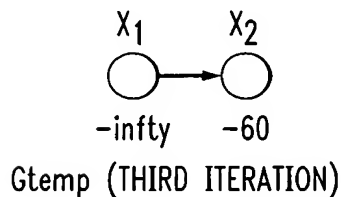
FIG. 4

The MaxIndependentSet CaRT Selection Algorithm

```
procedure MaxIndependentSet (T(X) , ē, G, neighborhood() )
Input:   n-attribute table T and n-vector of error tolerances ē;
         Bayesian network G on the set of attributes X and function
         neighborhood () defining the "predictive neighborhood" of a
         node Xᵢ in G (e.g., Π (X₁) or β (Xᵢ)).
Output:  A set of materialized (predicted) attributes Xₘₐₜ (Xₚᵣₑₐ = X -
         Xₘₐₜ) and a CaRT predictor PRED (Xᵢ)→Xᵢ for each Xᵢ ∈ Xₚᵣₑₐ.
begin
```

1. $X_{mat} := X$, $X_{pred} := \Phi$
2. $PRED(X_i) := \Phi$ for all $X_i \in X$, improve := true
3. while (improve $\neq$ false) do
4.   for each $X_i \in X_{mat}$
5.   mater_neighbors $(X_i) :=$
         $(X_{mat} \cap neighborhood(X_i)) \cup \{PRED(X) : X \in neighborhood$
         $(X_i) , X \in X_{pred}\} - \{X_i\}$
6. $M := BuildCaRT (Mater\_neighbors(X_i) \rightarrow X_i, e_i)$
7.   let $PRED(X_i) \subseteq mater\_neighbors(X_i)$ be the set of predictor attributes used in M
8. $cost\_change_i := 0$
9.   for each $X_j \in X_{pred}$ such that $X_i \in PRED(X_j)$
10. $NEW\_PRED_i(X_j) := PRED(X_j) - \{X_i\} \cup PRED(X_i)$
11. $M := BuildCaRT (NEW\_PRED_i(X_j) \rightarrow X_j, e_j$
12.   set $NEW\_PRED_i(X_j)$ to the (sub) set of predictor attributes used in M
13. $cost\_change_i := cost\_change_i + (PredCost(PRED(X_j) \rightarrow X_j) - PredCost(NEW\_PRED_i(X_j) \rightarrow X_j))$
14.   end
15.   end

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

5/8

FIG. 4 (cont)

16. build an undirected, node-weighted graph $G_{temp} = (X_{mat},$
    $E_{temp})$ on the current set of materialized
17. attributes X_{mat}, where:
18.   (a) $E_{temp} := \{ (X,Y) :$ for all pairs $X, Y \in X_{pred}\} \cup$
19. $\{ (X_i, Y) :$ for all $Y \in X_{mat}\}$
20.   (b) weight $(X_i) :=$ MaterCost (X_i) $-$PredCost (PRED(X_i)
    $\rightarrow X_i)$ +cost_change$_i$ for each $X_i \in X_{mat}$
21. $S :=$ FindWMIS (G_{temp}) /* select (approximate) maximum
    weight independent set in G_{temp}
22.   as "maximum–benefit" subset of
    predicted attributes */
23. if $(\Sigma_{X \in S}$ weight $(X) \leq 0)$ then improve := false
24. else/* update X_{mat}, X_{pred}, and the chosen CaRT predictors */
25.   for each $X_j \in X_{pred}$
26.   if (PRED $(X_j) \cap S = \{X_i\})$ then PRED $(X_j) :=$
    NEW_PRED$_i$ (X_j)
27.   end
28. $X_{mat} := X_{mat} - S$, $X_{pred} := X_{pred} \cup S$
29.   end
30. end /* while */
   end

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

6/8

FIG. 5

Algorithm for Estimating Lower Bound on Subtree Cost

```
procedure LowerBound      (N, e_i, b)
Input:   Leaf N for which lower bound on subtree cost is to be
         computed; error tolerance e_i for attribute X_i; bound b
         on the maximum number of internal nodes in subtree
         rooted at N.
Output: Lower bound L(N) on cost of subtree rooted at N.
begin
1.    for i := to r
2.         minOut [i, 0] :=i
3.    for J := 1 to b + 1
4.         minOut [0, j] :=0
5.    1 :=0
6.    for i := 1 to r
7.         while x_i - x_{1+1} > 2_{ei}
8.         1 :=1 = 1
9.    for j := 1 to b + 1
10.        minOut [i,j] := min {minOut[i - 1,j] + 1, minOut [1,j-1]
11. end
12. L(N) :=  ∞
13. for J := 0 to b
14.        L(N) :=min {L(N) , 2j + 1 + j log (|X_i|) + (j + 1 + minOut
           (r, j+1)) log (|dom(X_i)|)}
15. L(N) := min {L(N) , 2b + 3 + (b + 1) log (|X_i|) + (b +2) log
           (|dom(X_i) |) }
16. return L (N)
end
```
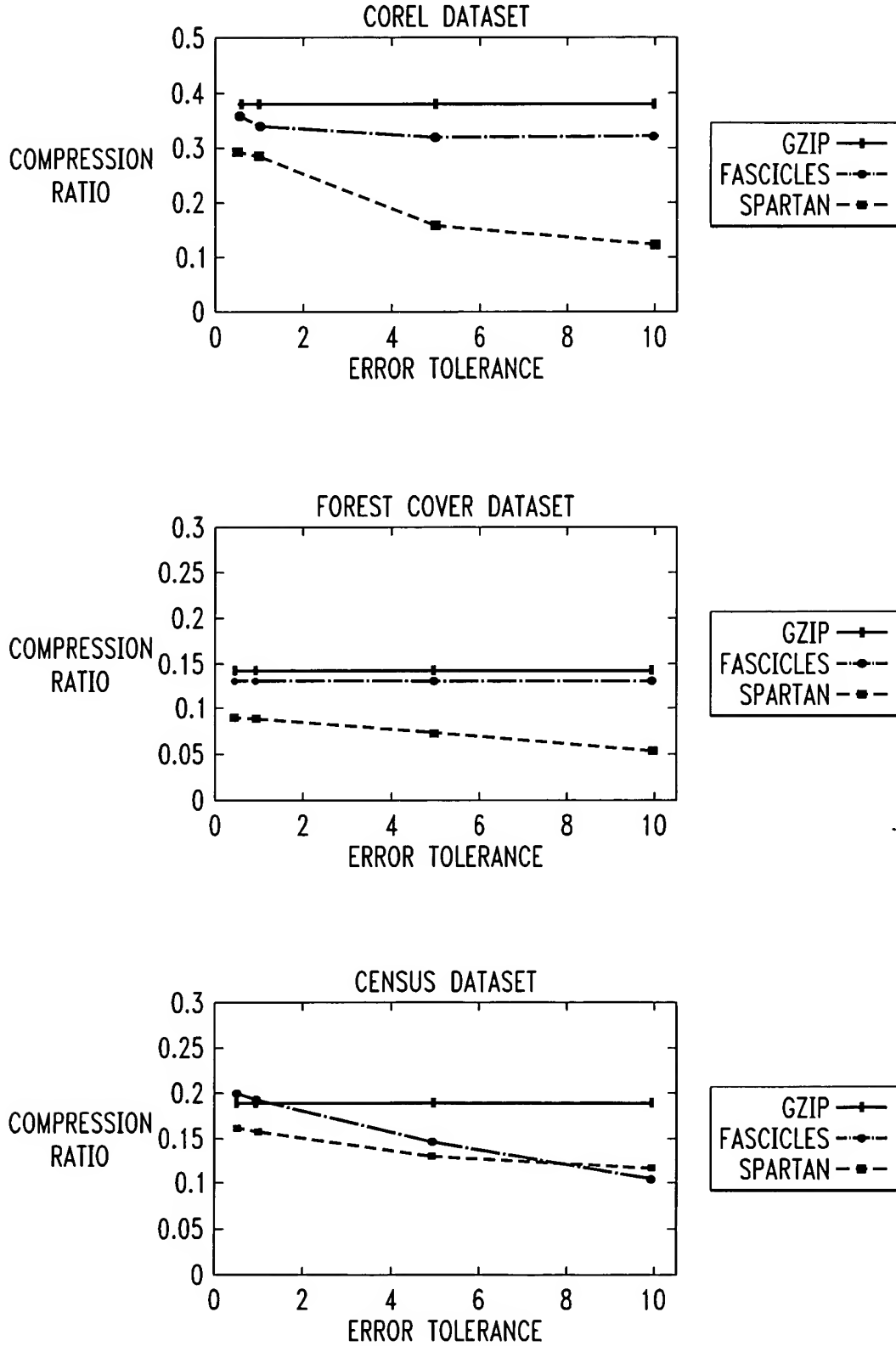
System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

7/8

FIG. 6

COREL DATASET

COMPRESSION RATIO

ERROR TOLERANCE

GZIP
FASCICLES
SPARTAN

FOREST COVER DATASET

COMPRESSION RATIO

ERROR TOLERANCE

GZIP
FASCICLES
SPARTAN

CENSUS DATASET

COMPRESSION RATIO

ERROR TOLERANCE

GZIP
FASCICLES
SPARTAN

System and Method for Compressing a Data Table
Babu 1-10-42
Serial No.: 10/033,199
Hitt, Gaines & Boisbrun, PC; (972) 480-8800

8/8

FIG. 7A

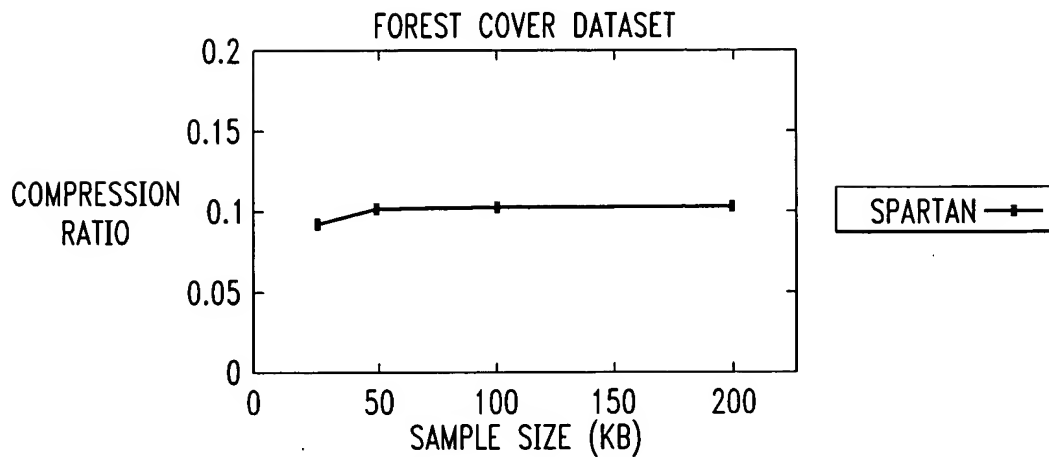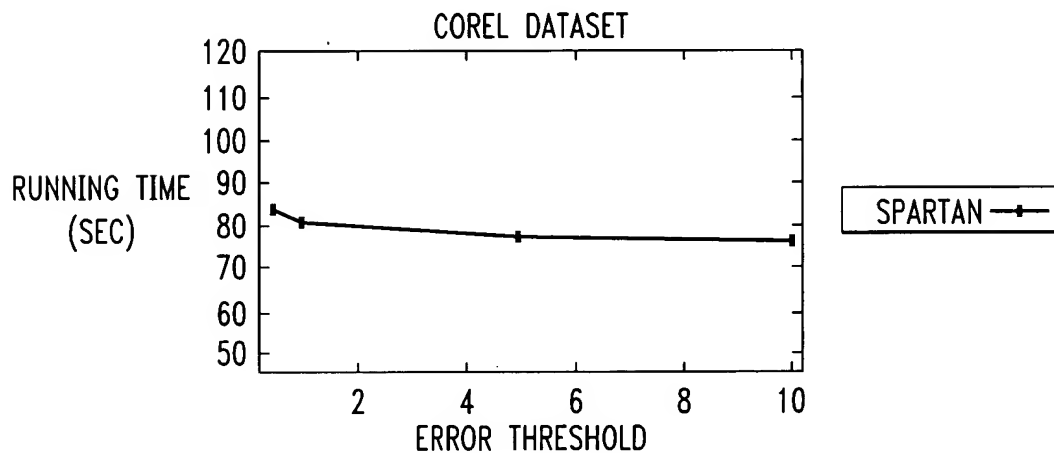FOREST COVER DATASET

FIG. 7B

COREL DATASET

FIG. 7C

COREL DATASET